



On the relationships between lumpability and filtering of finite stochastic systems

James Ledoux, Langford Barton White, Gary D. Brush

► To cite this version:

James Ledoux, Langford Barton White, Gary D. Brush. On the relationships between lumpability and filtering of finite stochastic systems. *Journal of Applied Probability*, 2008, 45 (3), pp.650-669. hal-00456111

HAL Id: hal-00456111

<https://hal.science/hal-00456111>

Submitted on 26 Aug 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the relationships between lumpability and filtering of finite stochastic systems

James Ledoux*, Langford B. White[†], Gary D. Bruse[‡]

27 August 2008

Abstract

The aim of this paper is to provide the conditions necessary to reduce the complexity of state filtering for finite stochastic systems (FSSs). A concept of lumpability for FSSs is introduced. This paper asserts that the unnormalised filter for a lumped FSS has linear dynamics. Two sufficient conditions for such a lumpability property to hold are discussed. It is shown that the first condition is also necessary for the lumped FSS to have a linear dynamics. Next, it is proven that the second condition allows the filter of the original FSS to be directly obtained from the filter for the lumped FSS. Finally, the paper generalises an earlier published result for the approximation of a general FSS by a lumpable one.

KEYWORDS: Optimal filtering, model reduction, hidden Markov chains, discrete markovian arrival process

AMS: 60J10; 93E11

1 Introduction

This paper deals with a class of homogeneous Markov chains (MC) $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ with finite state space $\mathcal{Y} \times \mathcal{X}$ and whose transition probabilities satisfy

$$\begin{aligned} P\{(Y_{t+1}, X_{t+1}) = (y, x) \mid (X_t, Y_t)\} &= P\{(Y_{t+1}, X_{t+1}) = (y, x) \mid X_t\} \\ &= D_y(X_t, x). \end{aligned} \tag{1}$$

*UMR-CNRS 6086 & Département de Mathématiques, BP 30179, Téléport 2, Boulevard Marie et Pierre Curie, F-86962 Futuroscope-Chasseneuil cedex, France

[†]School of Electrical and Electronic Engineering, University of Adelaide, SA 5005, Australia

[‡]Signals Analysis Discipline, C3I Division, Defence Science and Technology Organisation, P.O. Box 1500, Edinburgh, SA 5111, Australia

Such a Markov model is called a *Finite Stochastic System* (FSS) in [15, 1, 18] and is parametrised by the family of matrices $(D_y, y \in \mathcal{Y})$. A direct consequence of (1) is that $\{X_t\}_{t \in \mathbb{N}}$ is a Markov chain with transition probabilities

$$P(X_t, x) := \mathbb{P}\{X_{t+1} = x \mid X_t\} = \sum_{y \in \mathcal{Y}} D_y(X_t, x). \quad (2)$$

As a result, $\sum_{y \in \mathcal{Y}} D_y$ is a stochastic matrix.

Note that any discrete time Markovian arrival process $\{(N_t, X_t)\}_{t \in \mathbb{N}}$ (see [3]) defines an FSS setting $Y_t := N_t - N_{t-1}$ for $t \geq 1$. Using Bayes' formula, we have from (1) the general factorisation property:

$$D_y(X_t, x) = \mathbb{P}\{X_{t+1} = x \mid X_t\} \mathbb{P}\{Y_{t+1} = y \mid X_{t+1} = x, X_t\}. \quad (3)$$

Thus, the "output process" $\{Y_t\}_{t \in \mathbb{N}}$ may be thought of as generated by transitions of the Markov chain $\{X_t\}_{t \in \mathbb{N}}$. $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is called a *Hidden Markov Chain* (HMC) when the probabilities $\mathbb{P}\{Y_{t+1} = y \mid X_{t+1} = x, X_t\}$ in (3) do not depend on X_t and time t . If, for any t , the conditional distribution of Y_{t+1} given $X_{t+1} = x$ is denoted by $G(x, \cdot)$, this means that the transition probabilities (1) has the special factorisation property

$$D_y(X_t, x) = \mathbb{P}\{X_{t+1} = x \mid X_t\} \mathbb{P}\{Y_{t+1} = y \mid X_{t+1} = x\} = P(X_t, x) G(x, y). \quad (4)$$

Note that, in general, the distribution of Y_0 given X_0 is assumed to be $G(X_0, \cdot)$. It is well known that for an FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$, the process $\{((Y_t, (X_t, X_{t-1})))\}_{t \in \mathbb{N}^*}$ forms an HMC with state space $\mathcal{Y} \times \{(i, j) \in \mathcal{X} \times \mathcal{X} \mid P(i, j) > 0\}$. In this sense, the assumptions (1) and (4) on $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ are equivalent. However, only FSSs are considered in this paper for the following reasons. First, the factorisation properties (3) or (4) introduce no simplification in addressing the lumping and filtering problems. Moreover, the transformation of results on HMCs into results on FSS involves some notational and technical difficulties which are unhelpful to the reader. In contrast, any result on FSS can be applied to an HMC by replacing, everywhere, $D_y(X_t, x)$ by $P(X_t, x) G(x, y)$. Second, conceptual difficulties arise when you try to discuss some Markovian models widely used in stochastic modelling in the framework of HMCs [6].

For the rest of this paper $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ will denote an FSS. In this context, $\{Y_t\}_{t \in \mathbb{N}}$ and $\{X_t\}_{t \in \mathbb{N}}$ are called the *observed process* and the *state process* respectively. The aim of this paper is to provide a complexity reduction of the filtering for FSSs. More precisely, consider the *a posteriori* probabilities

$$\forall x \in \mathcal{X}, \quad \pi_t(x) := \mathbb{P}\{X_t = x \mid Y_t, \dots, Y_0\} \quad (5)$$

for $t \in \mathbb{N}$. Let π_t denotes the probability distribution $(\pi_t(x))_{x \in \mathcal{X}}$ on \mathcal{X} . $\{\pi_t\}_{t \in \mathbb{N}}$ is called the (state) *filter process* associated with $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$. We know that the filter process is the solution of the non-linear recursive equation (e.g. [13])

$$\pi_{t+1}(x) = \frac{\sum_{x_0 \in \mathcal{X}} \pi_t(x_0) D_{Y_{t+1}}(x_0, x)}{\sum_{x \in \mathcal{X}} \sum_{x_0 \in \mathcal{X}} \pi_t(x_0) D_{Y_{t+1}}(x_0, x)}, \quad (6)$$

where $D_{Y_{t+1}}$ is defined in a natural way. Another standard way of filtering is to use the unnormalised filter $\{\rho_t\}_{t \in \mathbb{N}} = \{(\rho_t(x, Y_t, \dots, Y_0))_{x \in \mathcal{X}}\}_{t \in \mathbb{N}}$ where, for every $t \in \mathbb{N}$ and $x \in \mathcal{X}$, $\rho_t(x, \cdot)$ is the positive measure on \mathcal{Y}^{t+1} defined by

$$\forall (y_t, \dots, y_0) \in \mathcal{Y}^{t+1}, \quad \rho_t(x, y_t, \dots, y_0) := \mathbb{P}\{X_t = x, Y_t = y_t, \dots, Y_0 = y_0\}. \quad (7)$$

We denote, for short, $\rho_t(x, Y_t, \dots, Y_0)$ by $\rho_t(x)$. The conditional probability (5) is obtained from

$$\pi_t(x) = \frac{\rho_t(x)}{\sum_{z \in \mathcal{X}} \rho_t(z)}. \quad (8)$$

The unnormalised filter has the main advantage of being the solution of the linear recursive equation

$$\rho_{t+1}(x) = \sum_{x_0 \in \mathcal{X}} \rho_t(x_0) D_{Y_{t+1}}(x_0, x). \quad (9)$$

We are interested in computing the filter associated with the function $\{g(X_t)\}_{t \in \mathbb{N}}$ of the Markov chain $\{X_t\}_{t \in \mathbb{N}}$ where $\text{card}(g(\mathcal{X})) < \text{card}(\mathcal{X})$, that is :

$$\forall w \in g(\mathcal{X}), \quad \hat{\pi}_t(w) := \mathbb{P}\{g(X_t) = w \mid Y_t, \dots, Y_0\}.$$

We introduce the unnormalised filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}} = \{(\hat{\rho}_t(w, Y_t, \dots, Y_0))_{w \in g(\mathcal{X})}\}_{t \in \mathbb{N}}$ associated with the lumped process $\{g(X_t)\}_{t \in \mathbb{N}}$, that is $\hat{\rho}_t(w, y_t, \dots, y_0)$ is defined, as in (7), by

$$\hat{\rho}_t(w, y_t, \dots, y_0) := \mathbb{P}\{g(X_t) = w, Y_t = y_t, \dots, Y_0 = y_0\}. \quad (10)$$

Using the same convention as in (8), we have

$$\hat{\pi}_t(w) = \frac{\hat{\rho}_t(w)}{\sum_{z \in g(\mathcal{X})} \hat{\rho}_t(z)}.$$

The problem here is that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is not an FSS in general, so that we can not use a linear recursive equation as in (9) for the computation of the unnormalised filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ of $\{g(X_t)\}_{t \in \mathbb{N}}$. The purpose of this paper is to propose conditions under which

$\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ could be derived using $\text{card}(g(\mathcal{X}))$ -dimensional matrix computations. A direct way is to look for conditions under which $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS, so that the filter $\{\hat{\pi}_t\}_{t \in \mathbb{N}}$ and the unnormalised filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ satisfy a recursive equation as in (6) and, respectively, in (9). A related problem was discussed in [21] for HMCs. In [21], a concept of strong lumpability for HMCs was defined. In that paper a general procedure for testing lumpability and deriving the associated lumped states was described. The present paper briefly discusses lumpability for FSSs using a somewhat more explicit relationship between lumpability of MCs and FSSs. Recalling that the focus is on the dynamics of the filter of lumped FSSs, the main contributions of the paper are:

- (1) The lumped filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ has linear dynamics irrespective of the probability distribution of (Y_0, X_0) if and only if the FSS is strongly lumpable w.r.t. function g .
- (2) A new condition is introduced for the unnormalised filter to have linear dynamics for some specific probability distributions of (Y_0, X_0) . Furthermore, this condition asserts that the filter $\{\pi_t\}_{t \in \mathbb{N}}$ can be directly computed from the lumped filter $\{\hat{\pi}_t\}_{t \in \mathbb{N}}$.

In Section 2, we revisit the basic results on lumpability of Markov chains in order to discuss the lumpability of FSSs in Section 3. The fact that the lumped filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ has linear dynamics if and only if the FSS is strongly lumpable is proven in Theorem 3.2. This new condition requiring $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ to have linear dynamics for a specific probability distribution of (Y_0, X_0) is introduced in Subsection 3.2, where we also show that when this property is true, the filter $\{\pi_t\}_{t \in \mathbb{N}}$ for the original FSS can be directly computed from the lumped filter $\{\hat{\pi}_t\}_{t \in \mathbb{N}}$. Section 4 discusses the problem of approximating an MC or an FSS by a strongly lumpable one, and proposes algorithms for computing such approximations.

2 Lumpable Markov Chains

Let $\{Z_t\}_{t \in \mathbb{N}}$ be a homogeneous Markov chain with state space $\mathcal{Z} = \{1, \dots, N\}$. Consider a function f from \mathcal{Z} into $f(\mathcal{Z}) = \{1, \dots, n\}$ with $n < N$. Such a map is called a *lumping map*. For notational convenience, f is assumed to be non-decreasing. This function defines a partition $\mathcal{Z}_i, i = 1, \dots, n$ of \mathcal{Z} where $\mathcal{Z}_i := f^{-1}(\{i\})$. The number of states in the subset \mathcal{Z}_i is denoted by N_i . We define the *lumping matrix* associated with this partition as the $N \times n$ matrix \mathbf{L} where $L(j, i) = 1$ when $j \in \mathcal{Z}_i$ and zero elsewhere. Next, we introduce the $n \times N$ -matrix \mathbf{U}

$$\mathbf{U} := (\mathbf{L}^\top \mathbf{L})^{-1} \mathbf{L}^\top = \text{diag}(1/N_i)^{-1} \mathbf{L}^\top. \quad (11)$$

The i -th row of \mathbf{U} is a N -dimensional probability vector which has precisely N_i non-zero elements, each with identical value $1/N_i$. Let \mathbf{Q} be the $N \times N$ -transition matrix of $\{Z_t\}_{t \in \mathbb{N}}$. $\mathbf{0}$ (resp. $\mathbf{1}$) will denote a matrix or vector with each entry equal to 0 (to 1), its dimension being defined by the context. Any vector is a row vector. In particular, the linear space $\text{Ker}(\mathbf{L})$ is defined by the set of vector $v \in \mathbb{R}^N$ such $v\mathbf{L} = \mathbf{0}$. \mathbf{I}_k will denote the $k \times k$ identity matrix.

The Markov chain $\{Z_t\}_{t \in \mathbb{N}}$ is said to be *lumpable* with respect to the function f and a specified initial distribution of Z_0 if $\{f(Z_t)\}_{t \in \mathbb{N}}$ is a homogeneous Markov chain. The Markov chain is said to be *strongly lumpable* with respect to f if it is lumpable with respect to f for every probability distribution of Z_0 .

Lemma 2.1 *The following statements are equivalent for $\{Z_t\}_{t \in \mathbb{N}}$ to be strongly lumpable.*

(a) *For all $w_1, w_2 \in f(\mathcal{Z})$,*

$$\mathbb{P}\{f(Z_{t+1}) = w_2 \mid Z_t = z_1\} = \mathbb{P}\{Z_{t+1} \in f^{-1}(w_2) \mid Z_t = z_1\} = \sum_{z_2 \in f^{-1}(w_2)} Q(z_1, z_2)$$

is independent of $z_1 \in f^{-1}(w_1)$; this conditional probability defines the transition probability from w_1 to w_2 for the Markov chain $\{f(Z_t)\}_{t \in \mathbb{N}}$;

(b) *the transition matrix \mathbf{Q} has the following block structure:*

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \cdots & \mathbf{Q}_{1n} \\ \vdots & & \vdots \\ \mathbf{Q}_{n1} & \cdots & \mathbf{Q}_{nn} \end{pmatrix}$$

where \mathbf{Q}_{ij} is a $N_i \times N_j$ -matrix which satisfies $\mathbf{Q}_{ij}\mathbf{1}^\top = q_{ij}\mathbf{1}^\top$ for some non-negative constant q_{ij} . The matrix $\hat{\mathbf{Q}} := (q_{ij})_{i,j=1,\dots,n}$ is stochastic;

(c) $\mathbf{QL} = \mathbf{LUQL}$;

(d) $\text{Ker}(\mathbf{L})\mathbf{Q} \subset \text{Ker}(\mathbf{L})$;

(e) $\text{Ker}(\mathbf{L}) \subset \text{Ker}(\mathbf{QL})$;

(f) *Let $\mathbf{V}^\top = \{v_1^\top, \dots, v_N^\top\}$, where the v_i are the right singular vectors of \mathbf{L}^\top . We have*

$$\mathbf{V}\mathbf{Q}\mathbf{V}^\top = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{B}_{22} \end{pmatrix} \quad (13)$$

where $\mathbf{B}_{11} \in \mathbb{R}_+^{n \times n}$.

In this case, $\{f(Z_t)\}_{t \in \mathbb{N}}$ is a Markov chain with transition matrix $\hat{Q} := UQL$.

Conditions (d) and (f) were proven in [21] for $n = 2$. Condition (d) for any $n \geq 2$ was derived in [10] from a very general criterion for a function of a Markov chain to be a Markov chain. That condition (f) for strong lumpability is valid for any $n \geq 2$ is proved in Appendix A. The other criteria are reviewed from [10, Theorem 10].

We now introduce another criterion for $\{f(Z_t)\}_{t \in \mathbb{N}}$ to be Markovian for some probability distribution of Z_0 . It was stated in [12] for finite state spaces and generalised in [16] for general state spaces. The following definition is from [10].

Definition 2.1 *Let L be the lumping matrix associated with lumping map f . A stochastic matrix Q is called a R-P matrix if there exist a $n \times N$ stochastic matrix Λ such that*

$$\Lambda L = I_n \quad \text{and} \quad \Lambda Q = \Lambda Q L \Lambda. \quad (14)$$

The following lemma states a known result on the lumpability of a Markov chain for which its transition matrix is a R-P matrix. In this case, note that the Markov property of $\{f(Z_t)\}_{t \in \mathbb{N}}$ depends on the probability distribution of Z_0 . This is a so-called *weak lumpability* condition [12].

Lemma 2.2 *Let $\{Z_t\}_{t \in \mathbb{N}}$ be a Markov chain with a R-P transition matrix Q . Then, the process $\{f(Z_t)\}_{t \in \mathbb{N}}$ is a Markov chain when the probability distribution of Z_0 is any convex combination of the n rows of matrix Λ . Moreover, its transition matrix is given by $\hat{Q} = \Lambda Q L$.*

When an MC has a R-P transition matrix with Λ defined as in (11) it is said to be exactly lumpable in [4]. See [10] for further properties of Markov chains with a R-P transition matrix.

Lumpability of MCs has been found to be relevant in various areas (e.g. see the recent papers [9, 14, 2, 20]). This is specially true in performance evaluation, where various modelling formalisms (e.g. Stochastic Automata Networks, Petri Nets and Algebra Processes) have been developed for model simplification. Every model specified by these formalisms has an underlying (continuous time) MC but, in general, with a very large state space. Thus, one objective is to avoid having to generate such a Markov graph. Therefore, the focus is on equivalence relations between basic objects of the formalism and on the development of efficient algorithms to aggregate equivalence classes. It is well known that some concepts of equivalence relation are directly connected to lumpability of the underlying MC (e.g. see [7, 11] for strong lumpability, [19] for R-P condition and references therein). But, in some sense, the lumpability of an MC is only used through the fundamental results given by [12] (statements (a)-(c) in Lemma 2.1 or Lemma 2.2). In this

way, the main contribution to the theory of lumpability of MCs is an efficient algorithm to find the optimal (strong) lumping map associated with a MC [5]. It will be clear from the next section that lumpability concepts for FSSs are directly related to lumpability for the bivariate MC. Thus, we do not contribute here to the theory of lumpability through new criteria, but we recall some equivalent forms which are not well-known. The aim of the paper is to investigate the connection between lumpability and dynamics of filters for FSSs in Section 3.

3 Lumpable finite stochastic systems and filtering

In the special case of an HMC, Spreij discussed the conditions under which the observed process $\{Y_t\}_{t \in \mathbb{N}}$ is a Markov chain [17]. The state process was assumed to be irreducible. This problem is solved in [10] under no particular assumption. The basic idea was to interpret the process $\{Y_t\}_{t \in \mathbb{N}}$ as the function $f(Y_t, X_t) = Y_t$ of the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ and to use criteria for lumpability of Markov chains. A similar idea, though not explicitly stated, was used in [21] to discuss the problem that we are interested in. Indeed, we are concerned here by the function $f(Y_t, X_t) = (Y_t, g(X_t))$ of the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ where g is some lumping map for the Markov chain $\{X_t\}_{t \in \mathbb{N}}$.

Definition 3.1 *An FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is said to be lumpable with respect to the lumping map g from \mathcal{X} into $g(\mathcal{X})$ if $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS for some probability distribution of (Y_0, X_0) . An FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is said to be strongly lumpable with respect to g if it is lumpable with respect to g for every probability distribution of (Y_0, X_0) .*

When an FSS is lumpable, note that both of the following conditions are satisfied: 1) $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is lumpable with respect to the function $f(y, x) = (y, g(x))$ from $\mathcal{Y} \times \mathcal{X}$ into $\mathcal{Y} \times g(\mathcal{X})$, and 2) the Markov chain $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is lumpable with respect to the function $f(y, x) = x$ from $\mathcal{Y} \times g(\mathcal{X})$ into $g(\mathcal{X})$ (the lumpability property above is relative to the same probability distribution of (Y_0, X_0)). However, the converse is not correct because the second condition only asserts that the conditional probabilities

$$P\{g(X_{t+1}) = w_{t+1} \mid Y_t = y_t, g(X_t) = w_t\}$$

are independent of y_t .

For the rest of this paper, $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is an FSS with finite state space $\mathcal{Y} \times \mathcal{X} = \{1, \dots, M\} \times \{1, \dots, N\}$ and g is a lumping map from \mathcal{X} into $g(\mathcal{X}) = \{1, \dots, n\}$. Recall that the transition matrix \mathbf{P} of the Markov chain $\{X_t\}_{t \in \mathbb{N}}$ is given by $\sum_{y \in \mathcal{Y}} \mathbf{D}_y$ (see (2)).

3.1 Strong lumpability for FSSs

Let us introduce our main result on the strong lumpability of FSSs.

Theorem 3.1 *Let \mathbf{L} be the lumping $N \times n$ -matrix associated with g . The FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is strongly lumpable with respect to g iff any of the following conditions are satisfied.*

(a) *For every $y \in \mathcal{Y}$, $\forall w_1, w_2 \in g(\mathcal{X})$,*

$$\mathbb{P}\{Y_{t+1} = y, g(X_{t+1}) = w_2 \mid X_t = x_1\} = \sum_{x_2 \in g^{-1}(w_2)} D_y(x_1, x_2) \quad (15)$$

is independent of $x_1 \in g^{-1}(w_1)$.

(b) *For each $y \in \mathcal{Y}$, the $N \times N$ -matrix \mathbf{D}_y has the following block structure:*

$$\mathbf{D}_y = \begin{pmatrix} \mathbf{E}_{11}(y) & \cdots & \mathbf{E}_{1n}(y) \\ \vdots & & \vdots \\ \mathbf{E}_{n1}(y) & \cdots & \mathbf{E}_{nn}(y) \end{pmatrix} \quad (16a)$$

where $\mathbf{E}_{ij}(y)$ is a $N_i \times N_j$ -matrix which satisfies

$$\mathbf{E}_{ij}(y) \mathbf{1}^\top = d_{ij}(y) \mathbf{1}^\top \quad (16b)$$

for some non-negative constant $d_{ij}(y)$. The matrices $\hat{\mathbf{D}}_y := (d_{ij}(y))_{i,j=1,\dots,n}$, $y \in \mathcal{Y}$ are such that $\sum_y \hat{\mathbf{D}}_y$ is a stochastic matrix.

(c) *For every $y \in \mathcal{Y}$,*

$$\mathbf{D}_y \mathbf{L} = \mathbf{L} \mathbf{U} \mathbf{D}_y \mathbf{L} \quad (17)$$

where $\mathbf{U} := (\mathbf{L}^\top \mathbf{L})^{-1} \mathbf{L}^\top$.

(d) *For every $y \in \mathcal{Y}$, $\text{Ker}(\mathbf{L}) \mathbf{D}_y \subset \text{Ker}(\mathbf{L})$.*

(e) *For every $y \in \mathcal{Y}$, $\text{Ker}(\mathbf{L}) \subset \text{Ker}(\mathbf{D}_y \mathbf{L})$.*

(f) *Set $\mathbf{V}^\top := \{\mathbf{v}_1^\top, \dots, \mathbf{v}_N^\top\}$, where the \mathbf{v}_i are the right singular vectors of \mathbf{L}^\top , we have for any $y \in \mathcal{Y}$*

$$\mathbf{V} \mathbf{D}_y \mathbf{V}^\top = \begin{pmatrix} \mathbf{B}_y & \mathbf{R}_y \\ \mathbf{0} & \mathbf{Z}_y \end{pmatrix} \quad (18)$$

where the \mathbf{B}_y are non-negative.

Moreover, $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS with characteristic matrices $(\hat{D}_y, y \in \mathcal{Y})$ given by

$$\hat{D}_y := U D_y L \quad \text{with } U := (L^\top L)^{-1} L^\top. \quad (19)$$

Proof. The main step is to verify that the FSS $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is strongly lumpable iff it is strongly lumpable as a Markov chain. The direct statement is obvious from the definition of an FSS. Now, assume that the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is strongly lumpable. From the first statement in Lemma 2.1, this is equivalent to the statement that for every $y \in \mathcal{Y}$, $\forall w_1, w_2 \in g(\mathcal{X})$,

$$P\{Y_{t+1} = y_1, g(X_{t+1}) = w_2 \mid Y_t = y_0, X_t = x_1\} \quad (20)$$

is independent of $x_1 \in g^{-1}(w_1)$. Since $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is an FSS, the probabilities above are independent of y_0 . Therefore, $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS for any probability distribution of (Y_0, X_0) and $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is strongly lumpable as an FSS.

Next, the equivalence of statements (a)–(f) is deduced from the six equivalent conditions in Lemma 2.1 for the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ to be strongly lumpable (see Appendix B). \square

Comment 1 *In this paper, we do not discuss the continuous-time counterpart of FSSs. But time discretisation by the standard uniformisation technique will provide the corresponding lumpability results for continuous-time FSSs. In this context, the entries of matrix D_y for an FSS must be interpreted as transition rates between states of the bivariate MC defining an FSS. Then, entries of matrix D_y are very similar to the labels of the derivation graph for a component in a Performance Evaluation Process Algebra (PEPA) model, this graph being the basis of generating the underlying continuous-time MC. In this framework, relation (15) may be thought of as the basic equality to define the concept of strong equivalence for PEPA. Note that such a relation is considered as a definition and it is shown that strong equivalence for PEPA implies strong lumpability of the underlying MC. In this way, this corresponds to the fact that Theorem 3.1 (a) (or basically Theorem 3.1 (b)) implies that the FSS is strongly lumpable with respect to g . We refer to [7, 11] for details and connections for other Stochastic Process Algebras.*

Comment 2 *When $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is an HMC then we know from (4) that*

$$D_y = P \text{diag}(\mathbf{G}(\cdot, y))$$

and the probability distribution of (Y_0, X_0) is given by

$$(\beta \text{diag}(\mathbf{G}(\cdot, 1)), \dots, \beta \text{diag}(\mathbf{G}(\cdot, M)))$$

where β is the probability distribution of X_0 . In general, $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is a Markov chain for every distribution of X_0 , is a weaker requirement than the Markov property of $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ for every probability distribution of (Y_0, X_0) .

Comment 3 In contrast to the Markov chain case, the “off-diagonal” condition

$$(C) \text{ “for every } y \in \mathcal{Y}, \forall w_1 \neq w_2 \in g(\mathcal{X}), \text{ the probability } \sum_{x_2 \in g^{-1}(w_2)} D_y(x_1, x_2) \text{ is independent of } x_1 \in g^{-1}(w_1)”$$

does not assert that the FSS is strongly lumpable (see Example 3.3). The definition in [21, page 2301] for an HMC to be (strongly) lumpable must be replaced by statement (a) in Theorem 3.1.

The main result of this subsection states that the strong lumpability condition provides the only way of ensuring that the unnormalised lumped filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ has linear dynamics.

Theorem 3.2 Let $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ be an FSS with characteristic matrices $(D_y, y \in \mathcal{Y})$. Consider the lumping map g with $N \times n$ -matrix \mathbf{L} and associated unnormalised lumped filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$. Then the following statements are equivalent:

(a) For any probability distribution of (Y_0, X_0) , $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ has the linear dynamics

$$\forall t \geq 0, \quad \hat{\rho}_{t+1} = \hat{\rho}_t \hat{\mathbf{D}}_{Y_{t+1}} \quad (21)$$

for some family of $n \times n$ -non-negative matrices $\hat{\mathbf{D}}_y, y \in \mathcal{Y}$ such that $\sum_{y \in \mathcal{Y}} \hat{\mathbf{D}}_y$ is a stochastic matrix;

(b) $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is a strongly lumpable FSS with respect to g ;

(c) $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is a strongly lumpable Markov chain with respect to g .

As a result, under any of the equivalent conditions of Theorem 3.1, the filter $\{\hat{\pi}_t\}_{t \in \mathbb{N}}$ associated with the lumped process $\{g(X_t)\}_{t \in \mathbb{N}}$ is given by

$$\hat{\pi}_t = \frac{\hat{\rho}_t}{\hat{\rho}_t \mathbf{1}^\top} \quad (22)$$

where the unnormalised filter $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ satisfies equation (21) with $\hat{\mathbf{D}}_y = \mathbf{U} \mathbf{D}_y \mathbf{L}, y \in \mathcal{Y}$.

Proof. The equivalence of statements (b) and (c) has been checked in the proof of Theorem 3.1. It is immediate from Theorem 3.1 that condition (b) implies statement (a). Indeed, if any of the four conditions in Theorem 3.1 is satisfied then $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS with characteristic matrices $(\hat{D}_y, y \in \mathcal{Y})$ whatever the probability distribution of (Y_0, X_0) . Thus the unnormalised filter satisfies a linear equation of the type (9), that is equation (21). A direct calculation may give some insight into this fact. Let $\{\rho_t\}_{t \in \mathbb{N}}$ be the unnormalised filter associated with the Markov chain $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$. It follows from (7) and (10) that $\{\rho_t\}_{t \in \mathbb{N}}$ and $\{\hat{\rho}_t\}_{t \in \mathbb{N}}$ are always related by:

$$\forall t \geq 0, \quad \hat{\rho}_t = \rho_t L. \quad (23)$$

Then, we can write for every $t \geq 0$

$$\begin{aligned} \hat{\rho}_{t+1} &= \rho_{t+1} L \\ &= \rho_t D_{Y_{t+1}} L && \text{from (9)} \\ &= \rho_t L U D_{Y_{t+1}} L && \text{from (17)} \\ &= \hat{\rho}_t U D_{Y_{t+1}} L && \text{from (23)} \\ &= \hat{\rho}_t \hat{D}_{Y_{t+1}} && \text{from (19)}. \end{aligned}$$

It remains to prove that statement (a) implies statement (b). This involves checking that condition (a) implies that, for any probability distribution of (Y_0, X_0) , the unnormalised filter $\{\rho_t\}_{t \in \mathbb{N}}$ satisfies the set of linear equations

$$\forall t \geq 0, \quad \rho_t (D_{Y_{t+1}} L - L \hat{D}_{Y_{t+1}}) = 0. \quad (24)$$

Indeed, we know from the algebraic manipulation above, that for any $t \geq 0$,

$$\hat{\rho}_{t+1} = \rho_t D_{Y_{t+1}} L$$

and it follows from condition (21) that

$$\hat{\rho}_{t+1} = \hat{\rho}_t \hat{D}_{Y_{t+1}} = \rho_t L \hat{D}_{Y_{t+1}}.$$

Combining the two representations of $\hat{\rho}_{t+1}$ we get (24).

Now, we can write

$$\rho_0 = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0}$$

where $\alpha_{y_0} = (P\{Y_0 = y_0, X_0 = x_0\})_{x_0 \in \mathcal{X}}$ and $1_{\{\cdot\}}$ is the indicator function. If we consider any probability distribution for (Y_0, X_0) it is clear that

$$\text{Span}(\alpha_{y_0}, y_0 \in \mathcal{Y}) = \mathbb{R}^N.$$

It follows from (24) that

$$\forall \alpha \in \mathbb{R}^N, \quad \alpha(D_{Y_1}L - L\widehat{D}_{Y_1}) = 0$$

or $\forall y_1 \in \mathcal{Y}$

$$D_{y_1}L - L\widehat{D}_{y_1} = 0.$$

Also, for any $y_1 \in \mathcal{Y}$, the equality $D_{y_1}L = L\widehat{D}_{y_1}$ implies that $UD_{y_1}L = UL\widehat{D}_{y_1} = \widehat{D}_{y_1}$. Then $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is a strongly lumpable FSS with respect to g from Theorem 3.1. \square

Example 3.3

Let us consider the following HMC (see (4)) with $N = 3, M = 2$, the lumping map $g(1) = 1, g(2) = g(3) = 2$ and matrices

$$P = \left(\begin{array}{c|cc} 1/2 & 1/3 & 1/6 \\ 1/4 & 1/2 & 1/4 \\ \hline 1/4 & 1/3 & 5/12 \end{array} \right) \quad G = \begin{pmatrix} 1/2 & 1/5 & 1/5 \\ 1/2 & 4/5 & 4/5 \end{pmatrix} \quad L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

This gives an FSS with characteristic matrices $D_1 = P \text{diag}(G(\cdot, 1))$, $D_2 = P \text{diag}(G(\cdot, 2))$:

$$D_1 = \left(\begin{array}{c|cc} 1/4 & 1/15 & 1/30 \\ 1/8 & 1/10 & 1/20 \\ \hline 1/8 & 1/15 & 1/12 \end{array} \right) \quad D_2 = \left(\begin{array}{c|cc} 1/4 & 4/15 & 2/15 \\ 1/8 & 2/5 & 1/5 \\ \hline 1/8 & 4/15 & 1/3 \end{array} \right)$$

and we have

$$PL = \left(\begin{array}{c|cc} 1/2 & 1/2 \\ 1/4 & 3/4 \\ \hline 1/4 & 3/4 \end{array} \right) \quad D_1L = \left(\begin{array}{c|cc} 1/4 & 1/10 \\ 1/8 & 3/20 \\ \hline 1/8 & 3/20 \end{array} \right) \quad D_2L = \left(\begin{array}{c|cc} 1/4 & 2/5 \\ 1/8 & 3/5 \\ \hline 1/8 & 3/5 \end{array} \right).$$

Then, statement (a) is met and $\{(Y_t, g(X_t))\}_{t \geq 0}$ is an FSS with characteristic matrices

$$\widehat{D}_1 = \begin{pmatrix} 1/4 & 1/10 \\ 1/8 & 3/20 \end{pmatrix} \quad \widehat{D}_2 = \begin{pmatrix} 1/4 & 2/5 \\ 1/8 & 3/5 \end{pmatrix}.$$

Now, we deduce from Theorem 3.2 that the filters can be computed from observations (Y_0, \dots, Y_t) as follows (see also (17)): for any probability vector α on \mathcal{X} ,

$$\rho_t = \alpha \text{diag}(G(\cdot, Y_0))D_{Y_1} \cdots D_{Y_t} \quad \text{and} \quad \widehat{\rho}_t = \rho_t L = \alpha \text{diag}(G(\cdot, Y_0))L\widehat{D}_{Y_1} \cdots \widehat{D}_{Y_t} \\ = \widehat{\rho}_0 \widehat{D}_{Y_1} \cdots \widehat{D}_{Y_t}$$

$$\pi_t = \frac{\alpha \text{diag}(G(\cdot, Y_0))D_{Y_1} \cdots D_{Y_t}}{\alpha \text{diag}(G(\cdot, Y_0))D_{Y_1} \cdots D_{Y_t} \mathbf{1}^\top} \quad \text{and} \quad \widehat{\pi}_t = \frac{\widehat{\rho}_0 \widehat{D}_{Y_1} \cdots \widehat{D}_{Y_t}}{\widehat{\rho}_0 \widehat{D}_{Y_1} \cdots \widehat{D}_{Y_t} \mathbf{1}^\top}.$$

In fact, Theorem 3.2 asserts that the previous equalities hold for the FSS with characteristic matrices $(\mathbf{D}_1, \mathbf{D}_2)$ deduced from this HMC, and are valid for every probability distribution of (Y_0, X_0) (see related Comment 2)

Now, replace matrix \mathbf{G} in the previous model by:

$$\mathbf{G}_1 = \begin{pmatrix} 1/2 & 1/4 & 1/3 \\ 1/2 & 3/4 & 2/3 \end{pmatrix}.$$

A direct calculation gives:

$$\begin{aligned} & \mathbb{P}\{Y_2 = 1, g(X_2) = 2 \mid Y_1 = 1, g(X_2) = 2, Y_0 = 2, X_0 = 1\} = 77/360 \\ & \text{and } \mathbb{P}\{Y_2 = 1, g(X_2) = 2 \mid Y_1 = 2, g(X_2) = 2, Y_0 = 2, X_0 = 1\} = 77/936 \end{aligned}$$

so that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is not an FSS (and not an HMC as well). Note that $\{g(X_t)\}_{t \in \mathbb{N}}$ is an MC for every probability distribution of X_0 from Lemma 2.1 (see the form of matrix $\mathbf{P}\mathbf{L}$ above). Finally, it is easily checked that

$$\mathbf{D}_1 \mathbf{L} = \left(\begin{array}{c|c} 1/4 & 5/36 \\ \hline 1/8 & 5/24 \\ 1/8 & 2/9 \end{array} \right) \quad \mathbf{D}_2 \mathbf{L} = \left(\begin{array}{c|c} 1/4 & 13/36 \\ \hline 1/8 & 13/24 \\ 1/8 & 19/36 \end{array} \right).$$

Therefore condition (C) in Comment 3 is satisfied although the FSS is not lumpable.

3.2 Rogers-Pitman's condition for FSSs

In this part, we deal with a criterion which will be deduced from the Rogers-Pitman's criterion for the lumpability of MCs. In order to ease the exposition, $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ will be thought of as an univariate Markov chain $\{Z_t\}_{t \in \mathbb{N}}$ with state space $\mathcal{Z} := \{1, \dots, NM\}$ using a lexicographic ordering of the elements of $\mathcal{Y} \times \mathcal{X}$:

$$x \in \mathcal{X}, y \in \mathcal{Y}, \quad Z_t = (y - 1)N + x \iff (Y_t, X_t) = (y, x).$$

When $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ is an FSS, its $NM \times NM$ -transition matrix has the form

$$\mathbf{Q} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_M \\ \vdots & \vdots & & \vdots \\ \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_M \end{pmatrix} \quad (25)$$

or in a compact form

$$\mathbf{Q} = (\mathbf{1}_M^\top \otimes \mathbf{I}_N) (\mathbf{D}_1 \cdots \mathbf{D}_M) \quad (26)$$

using the Kronecker product \otimes of matrices. The probability distribution of Z_0 is the NM -vector α obtained by listing the components of the $M \times N$ -matrix $(P\{Y_0 = y_0, X_0 = x_0\})_{y_0 \in \mathcal{Y}, x_0 \in \mathcal{X}}$ using the lexicographic ordering. Next, the process $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is associated with the function $\{f(Z_t)\}_{t \in \mathbb{N}}$ of $\{Z_t\}_{t \in \mathbb{N}}$ where f is defined by

$$\forall y \in \mathcal{Y}, \forall x \in \mathcal{X}, \quad f((y-1)N + x) := (y-1)n + g(x). \quad (27)$$

Then, the corresponding lumping $MN \times Mn$ -matrix is

$$\mathbf{L}_f = \begin{pmatrix} \mathbf{L}_g & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{L}_g \end{pmatrix} \quad (28)$$

where $\mathbf{0} \in \mathbb{R}^{N \times n}$ and \mathbf{L}_g is the lumping $N \times n$ -matrix associated with g . In a compact form, $\mathbf{L}_f = \mathbf{I}_M \otimes \mathbf{L}_g$.

It is clear that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is a Markov chain (an FSS) for $(P\{Y_0 = y_0, X_0 = x_0\})_{y_0 \in \mathcal{Y}, x_0 \in \mathcal{X}}$ iff $\{f(Z_t)\}_{t \in \mathbb{N}}$ is a Markov chain (an FSS) with α as probability distribution of Z_0 . Now, we state a condition for $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ to be an FSS for specific probability distributions of (Y_0, X_0) .

Theorem 3.4 *Let Λ_g be a stochastic $n \times N$ -matrix such that*

$$\Lambda_g \mathbf{L}_g = \mathbf{I}_n. \quad (29)$$

Let us introduce the subset \mathcal{P} of probability distributions ν on $\mathcal{Y} \times \mathcal{X}$ such that

$$\nu(y, x) = \begin{cases} \Lambda_g(w, x) & \text{if } g(x) = w \\ 0 & \text{otherwise.} \end{cases} \quad (30)$$

We denote the set of convex combinations of elements of \mathcal{P} by $\text{Conv}(\mathcal{P})$.

If we have

$$\Lambda_g \mathbf{D}_y = \Lambda_g \mathbf{D}_y \mathbf{L}_g \Lambda_g \quad (31)$$

for every $y \in \mathcal{Y}$, then $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS for every probability distribution of (Y_0, X_0) in $\text{Conv}(\mathcal{P})$. Its characteristic matrices $(\hat{\mathbf{D}}_y, y \in \mathcal{Y})$ are given by

$$\forall y \in \mathcal{Y}, \quad \hat{\mathbf{D}}_y := \Lambda_g \mathbf{D}_y \mathbf{L}_g. \quad (32)$$

Proof. The first step consists in checking that conditions (31) assert that the process $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an MC for every probability distribution of (Y_0, X_0) in $\text{Conv}(\mathcal{P})$. Using the univariate framework proposed in the beginning of this paragraph, we must check that the process $\{f(Z_t)\}_{t \in \mathbb{N}}$ satisfies Definition 2.1. Let us introduce the stochastic $Mn \times MN$ block diagonal matrix

$$\Lambda_f = I_M \otimes \Lambda_g. \quad (33)$$

Observe that $\Lambda_f L_f = I_{Mn}$ from relation (29). Each row of the matrix Λ_f corresponds to an unique probability distribution in the set \mathcal{P} . Next using a block-decomposition of each of matrices, it is easily seen from (26), (28) and (33) that

$$\Lambda_f Q = \begin{pmatrix} \Lambda_g D_1 & \Lambda_g D_2 & \cdots & \Lambda_g D_M \\ \vdots & \vdots & & \vdots \\ \Lambda_g D_1 & \Lambda_g D_2 & \cdots & \Lambda_g D_M \end{pmatrix}$$

and

$$\Lambda_f Q L_f = \begin{pmatrix} \Lambda_g D_1 L_g & \Lambda_g D_2 L_g & \cdots & \Lambda_g D_M L_g \\ \vdots & \vdots & & \vdots \\ \Lambda_g D_1 L_g & \Lambda_g D_2 L_g & \cdots & \Lambda_g D_M L_g \end{pmatrix}. \quad (34)$$

Therefore, $\Lambda_f Q = \Lambda_f Q L_f \Lambda_f$ if and only if we have for every $y \in \mathcal{Y}$:

$$\Lambda_g D_y = \Lambda_g D_y L_g \Lambda_g.$$

This last equality is just equation (31). From Lemma 2.2, we deduce that $\{f(Z_t)\}_{t \in \mathbb{N}}$ is an MC for every probability distribution of Z_0 which is a convex combination of the rows of Λ_f . Its transition matrix \hat{Q} has the form

$$\hat{Q} := \Lambda_f Q L_f = \begin{pmatrix} \hat{D}_1 & \hat{D}_2 & \cdots & \hat{D}_M \\ \vdots & \vdots & & \vdots \\ \hat{D}_1 & \hat{D}_2 & \cdots & \hat{D}_M \end{pmatrix}$$

where $\hat{D}_y := \Lambda_g D_y L_g$. Thus from the structure of the matrix \hat{Q} it is clear that $\{f(Z_t)\}_{t \in \mathbb{N}}$ is an FSS (e.g. see (25)). \square

Comment 4 When $\Lambda_g = (L_g^\top L_g)^{-1} L_g^\top$, conditions (31) in Theorem 3.4 reads as follows: for every $y \in \mathcal{Y}$, $\forall w_1, w_2 \in g(\mathcal{X})$,

$$\sum_{x_1 \in g^{-1}(w_1)} D_y(x_1, x_2) \quad (35)$$

is independent of $x_2 \in g^{-1}(w_2)$. This relation is the exact lumpability condition discussed in [4] and corresponds to backward simulation (see [19] and the references therein). This appears as a “dual condition” of that reported in Theorem 3.1 (see (15)). The duality between strong lumpability and Rogers-Pitman’s condition is made explicit in the context of Markov chains in [10, Paragraph 2.2.4.].

The next result shows the interest in Rogers-Pitman’s condition for filtering an FSS.

Theorem 3.5 *Using the same notation as in Theorem 3.4, Let $\{(Y_t, X_t)\}_{t \in \mathbb{N}}$ be an FSS with characteristic matrices $(\mathbf{D}_y, y \in \mathcal{Y})$. Then, the two following statements are equivalent:*

(a) *for any $y \in \mathcal{Y}$,*

$$\Lambda_g \mathbf{D}_y = \Lambda_g \mathbf{D}_y \mathbf{L}_g \Lambda_g; \quad (36)$$

(b) *for any probability distribution of (Y_0, X_0) in $\text{Conv}(\mathcal{P})$, for any $t \geq 0$,*

$$\hat{\rho}_{t+1} = \hat{\rho}_t \hat{\mathbf{D}}_{Y_{t+1}} \quad \text{and} \quad \rho_t = \hat{\rho}_t \Lambda_g \quad (37)$$

for some family of $n \times n$ -matrices $(\hat{\mathbf{D}}_y, y \in \mathcal{Y})$ such that $\sum_y \hat{\mathbf{D}}_y$ is stochastic.

Proof. First we prove that (a) \implies (b). The expressions for ρ_0 and $\hat{\rho}_0$ in connection with our choice of probability distribution of (Y_0, X_0) in $\text{Conv}(\mathcal{P})$, can be written from (7), (23) and (29) as:

$$\rho_0 = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0} \Lambda_g \quad \text{and} \quad \hat{\rho}_0 = \rho_0 \mathbf{L}_g = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0} \quad (38)$$

with

$$\forall y_0 \in \mathcal{Y}, \alpha_{y_0} \in \mathbb{R}_+^n \text{ and } \sum_{y_0 \in \mathcal{Y}} \alpha_{y_0} \mathbf{1}^\top = 1.$$

The leftmost equality in (37) follows from the fact that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS with characteristic matrices $(\hat{\mathbf{D}}_y, y \in \mathcal{Y})$ for every distribution of (Y_0, X_0) in $\text{Conv}(\mathcal{P})$. Proof of the rightmost equality is by induction on t . For $t = 0$, the relation is deduced from (38):

$$\rho_0 = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0} \Lambda_g = \left(\sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0} \right) \Lambda_g = \hat{\rho}_0 \Lambda_g.$$

Assume that $\rho_t = \hat{\rho}_t \Lambda_g$. Then,

$$\begin{aligned}\rho_{t+1} &= \rho_t D_{Y_{t+1}} \\ &= \hat{\rho}_t \Lambda_g D_{Y_{t+1}} \\ &= \hat{\rho}_t \hat{D}_{Y_{t+1}} \Lambda_g \quad \text{from (36) and (32)} \\ &= \hat{\rho}_{t+1} \Lambda_g\end{aligned}$$

from the leftmost equality in (37).

Conversely, assume that statement (b) holds. Then, from relations (37) : for every $t \in \mathbb{N}$

$$\rho_{t+1} = \rho_t D_{Y_{t+1}} = \hat{\rho}_t \Lambda_g D_{Y_{t+1}} \quad \text{and} \quad \rho_{t+1} = \hat{\rho}_{t+1} \Lambda_g = \hat{\rho}_t \hat{D}_{Y_{t+1}} \Lambda_g.$$

In combining the two representations of ρ_{t+1} for any $t \in \mathbb{N}$ we obtain

$$\hat{\rho}_t (\Lambda_g D_{Y_{t+1}} - \hat{D}_{Y_{t+1}} \Lambda_g) = 0$$

or $\forall y \in \mathcal{Y}$

$$\hat{\rho}_t (\Lambda_g D_y - \hat{D}_y \Lambda_g) = 0.$$

In particular, we have for $t = 0$:

$$\forall y \in \mathcal{Y}, \quad \hat{\rho}_0 (\Lambda_g D_y - \hat{D}_y \Lambda_g) = 0. \quad (39)$$

Now, we write from (38)

$$\hat{\rho}_0 = \sum_{y_0 \in \mathcal{Y}} 1_{\{Y_0=y_0\}} \alpha_{y_0}$$

where $\alpha_{y_0} \in \mathbb{R}_+^n$ and $\sum_{y_0} \alpha_{y_0} \mathbf{1}^\top = 1$. Now, it is clear that $\text{Span}(\alpha_{y_0}, y_0 \in \mathcal{Y}) = \mathbb{R}^n$. Then, we deduce from (39) that

$$\forall y \in \mathcal{Y}, \forall \alpha \in \mathbb{R}^n, \quad \alpha (\Lambda_g D_y L_g - \hat{D}_y \Lambda_g) = 0$$

so that

$$\forall y \in \mathcal{Y}, \quad \Lambda_g D_y L_g = \hat{D}_y \Lambda_g.$$

Next, for any $y \in \mathcal{Y}$, the equality $\Lambda_g D_y = \hat{D}_y \Lambda_g$ implies that $\Lambda_g D_y L_g = \hat{D}_y \Lambda_g L_g = \hat{D}_y$. Then statement (a) is met. \square

Example 3.6

Let us illustrate the results by an example. We consider the same lumping map as in Example 3.3 for the FSS with characteristic matrices (D_1, D_2) :

$$D_1 := \left(\begin{array}{c|cc} 1/4 & 1/20 & 1/20 \\ \hline 5/24 & 1/5 & 2/15 \\ \hline 1/8 & 2/15 & 1/5 \end{array} \right) \quad D_2 := \left(\begin{array}{c|cc} 1/4 & 1/5 & 1/5 \\ \hline 1/8 & 2/15 & 1/5 \\ \hline 5/24 & 1/5 & 2/15 \end{array} \right).$$

Since $D_1(2, 1) \neq D_1(3, 1)$, the FSS is not strongly lumpable from Theorem 3.1. Note that the transition matrix of the Markov chain $\{X_t\}_{t \in \mathbb{N}}$

$$P = D_1 + D_2 = \left(\begin{array}{c|cc} 1/2 & 1/4 & 1/4 \\ \hline 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{array} \right)$$

is strongly lumpable from Lemma 2.1. We introduce the following matrix Λ_g :

$$\Lambda_g := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix}.$$

Note that P is also a R-P matrix for matrix Λ_g from Lemma 2.2.

Next, we obtain

$$\widehat{D}_1 := \Lambda_g D_1 L_g = \begin{pmatrix} 1/4 & 1/10 \\ 1/6 & 1/3 \end{pmatrix} \quad \widehat{D}_2 := \Lambda_g D_2 L_g = \begin{pmatrix} 1/4 & 2/5 \\ 1/6 & 1/3 \end{pmatrix}$$

It is easily checked that $\Lambda_g D_i = \widehat{D}_i \Lambda_g$ for $i = 1, 2$, so that $\{(Y_t, g(X_t))\}_{t \in \mathbb{N}}$ is an FSS with characteristic matrices $(\widehat{D}_1, \widehat{D}_2)$ for every probability distribution of (Y_0, X_0) of the form $\alpha \Lambda_g = \alpha(1)(1, 0, 0) + (1 - \alpha(1))(0, 1/2, 1/2)$ where α is any stochastic vector on $g(\mathcal{X})$. For any sequence of observations (Y_0, \dots, Y_t) , the filters are given from Theorem 3.4 as follows. For every stochastic 2-dimensional vector α ,

$$\widehat{\rho}_t = \alpha \widehat{D}_{Y_0} \cdots \widehat{D}_{Y_t} \quad \rho_t = \widehat{\rho}_t \Lambda_g = \alpha \widehat{D}_{Y_0} \cdots \widehat{D}_{Y_t} \Lambda_g = \alpha \Lambda_g D_{Y_0} \cdots D_{Y_t}.$$

4 Approximation by strongly lumpable models

In [21] a procedure was derived that yielded a 2-lumpable approximation to a given MC, that is an approximation of a given MC by a strongly lumpable MC with two lumps ($n = 2$). By extension, 2-lumpable approximations to HMCs were obtained. If this procedure is recursively repeated on each of the two lumpings until no more lumpings are possible then the approximate n -lumpable HMC will be determined and the filters for the approximate states computed. However, as the number of lumpings, n , approaches the number of states, N , in the original HMC, the question arises as to benefits of determining any more lumpings. Also, there is no obvious procedure to obtain an optimal approximation when the lumping mapping is unknown. Note that for large state spaces, a crucial question is to search for a lumping map. But this point is beyond the scope of the present paper. We mention in passing, that there is an algorithm with $O(k \log N)$ time complexity (k is the

number of non-zero transition probabilities) that computes the optimal lumping map for any MC [5] (which is optimal in terms of reduction of state space complexity). However, it remains unknown as to how to find a best lumping in some approximate sense when the MC is not lumpable.

In Subsection 4.1, we generalise the problem presented in [21] for MCs, for $n > 2$, i.e. we assume we have knowledge of the lumping map g for the process, and wish to determine the closest approximation (in the Frobenius norm sense) of the transition probability matrix on the assumed lumping. The results are generalised to FSSs using a process outlined in Subsection 4.2.

4.1 Approximation of MCs

Let

$$\mathcal{T}_N(\mathbf{L}) = \{ \Psi \in \mathbb{R}^{N \times N} : \text{Ker}(\mathbf{L})\Psi \subset \text{Ker}(\mathbf{L}) \} .$$

then $\Psi \in \mathcal{T}_N(\mathbf{L})$ iff

$$\Psi = \mathbf{V}^\top \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{0} & \mathbf{B}_{22} \end{pmatrix} \mathbf{V}, \quad (40)$$

where $\mathbf{B}_{11} \in \mathbb{R}^{n \times n}$, $\mathbf{B}_{12} \in \mathbb{R}^{n \times (N-n)}$, $\mathbf{B}_{22} \in \mathbb{R}^{(N-n) \times (N-n)}$, $\mathbf{V}^\top = [\mathbf{v}_1^\top, \dots, \mathbf{v}_N^\top]$ and \mathbf{v}_i s are the right singular vectors of \mathbf{L}^\top , arranged such that $\text{Ker}(\mathbf{L}) = \text{Span}\{\mathbf{v}_{n+1}, \dots, \mathbf{v}_N\}$. It can be seen directly from (40) that $\mathcal{T}_N(\mathbf{L})$ has dimension $N^2 - n(N-n)$. Note that we do not restrict elements of $\mathcal{T}_N(\mathbf{L})$ to be stochastic, or even non-negative matrices. The proof of this result follows in the same way as in the proof of Lemma 1 given in Appendix A.

Lemma 4.1 *The orthogonal projection onto the subspace $\mathcal{T}_N(\mathbf{L})$ of $\mathbb{R}^{N \times N}$ is given by*

$$\Pi(\mathbf{Q}) = \mathbf{Q} - (\mathbf{I}_N - \Phi) \mathbf{Q} \Phi \quad \text{where } \Phi = \mathbf{V}^\top \begin{pmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}. \quad (41)$$

Proof. Let $\mathbf{Q} \in \mathbb{R}^{N \times N}$, then we seek a matrix $\Psi \in \mathcal{T}_N(\mathbf{L})$ which minimises $\|\mathbf{Q} - \Psi\|$. Let $\mathbf{B} = \mathbf{V}\Psi\mathbf{V}^\top$, and $\mathbf{C} = \mathbf{V}\mathbf{Q}\mathbf{V}^\top$. We know that \mathbf{B} has the form of the block matrix in (40). Partition \mathbf{C} commensurately with \mathbf{B} , then clearly $\|\mathbf{Q} - \Psi\| = \|\mathbf{C} - \mathbf{B}\|$ is minimised by taking

$$\Psi = \mathbf{V}^\top \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{0} & \mathbf{C}_{22} \end{bmatrix} \mathbf{V}.$$

It is simple to verify that indeed, Ψ has the form (41). □

If \mathbf{Q} is a stochastic matrix, then in general, $\Pi(\mathbf{Q})$ will not be. In order to obtain a stochastic matrix corresponding to the transition probability matrix of a strongly lumpable MC, we employ the method of alternating convex projections [23]. Apart from the above subspace constraint, we also need to force non-negativity of the elements of the approximant, and the property that the row sums are unity. We do this applying successively the following two projection operators to $\Pi(\mathbf{Q})$. Let us consider the subset of $\mathbb{R}^{N \times N}$ consisting of non-negative matrices. It can be shown that the (convex) projection of any $\mathbf{M} \in \mathbb{R}^{N \times N}$ onto this subset is

$$\Pi_+(\mathbf{M})(i, j) = \begin{cases} M(i, j) & \text{if } M(i, j) \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Now consider the affine subspace of $\mathbb{R}^{N \times N}$ consisting of matrices with all row sums unity. It can be shown that the (affine) projection of $\mathbf{M} \in \mathbb{R}^{N \times N}$ onto this subspace is given by

$$\Pi_1(\mathbf{M}) = \mathbf{M} \left(\mathbf{I} - \frac{\mathbf{1}^\top \mathbf{1}}{N} \right) + \frac{\mathbf{1}^\top \mathbf{1}}{N}. \quad (42)$$

The algorithm proposed in [21] consists of repeatedly applying each of the three projection operators above to the given matrix \mathbf{Q} until convergence is noted. It is known that the algorithm will converge to a stochastic matrix being the transition probability matrix of a strongly lumpable Markov chain, although the optimality of the approximation has not been demonstrated.

Let us illustrate how the procedure works using two examples. The first is didactic, whilst the second illustrates the convergence behaviour of the algorithm in a more realistic case. Finally, we compare the filtering performance for the aggregated states of the optimal filter, and that filter obtained from a lumpable approximation to the model.

Example 4.1

Consider a MC with $N = 3$ states, which we wish to lump into $n = 2$ groups under the lumping matrix

$$\mathbf{L} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.$$

We start with a matrix \mathbf{Q}_0 which is constructed by randomly perturbing the transition probability matrix \mathbf{Q} of a MC lumpable under \mathbf{L} (with appropriate renormalisation)

$$\mathbf{Q} = \left(\begin{array}{c|cc} 1/2 & 1/4 & 1/4 \\ \hline 1/3 & 1/3 & 1/3 \\ \hline 1/3 & 1/3 & 1/3 \end{array} \right).$$

The initial iterate is determined by adding a random matrix with independent and identically distributed entries uniform on $[-0.1, 0.1]$ and renormalising giving

$$\mathbf{Q}_0 = \begin{pmatrix} 0.5471 & 0.2292 & 0.2237 \\ 0.3012 & 0.4434 & 0.2554 \\ 0.3117 & 0.3390 & 0.349 \end{pmatrix}.$$

The algorithm terminates after one iteration giving the approximate transition matrix

$$\mathbf{Q}_1 = \begin{pmatrix} 0.5471 & 0.2292 & 0.2237 \\ 0.3064 & 0.4408 & 0.2528 \\ 0.3064 & 0.3416 & 0.3520 \end{pmatrix},$$

corresponding to the transition matrix of an MC lumpable under \mathbf{L} . Note however that the algorithm does not recover \mathbf{Q} . Note that $\|\mathbf{Q}_1 - \mathbf{Q}\| = 0.1525$, and $\|\mathbf{Q}_1 - \mathbf{Q}_0\| = 0.091$, so indeed the solution is closer to the initial iterate than to \mathbf{Q} .

The next example has $N = 100$ and $n = 2$. The sizes of the groups were $N_1 = 40$ and $N_2 = 60$. We ran 1000 independent realisations with random initial transition matrices. The number of iterations until the relative error between successive iterates dropped to 10^{-8} was determined, and a histogram of the results is shown in Figure 1. These results show, and it supported by other experiments, that the algorithm converges in only a few steps, even for relatively large problems.

4.2 Approximation of FSSs

An FSS is characterised by a finite set $(\mathbf{D}_y : y \in \mathcal{Y})$ of sub-stochastic matrices. The above approximation procedure can also be applied to FSSs with one modification. The lumpability and non-negativity constraints can be applied to each \mathbf{D}_y independently, however the stochasticity constraint needs to be applied to the $N \times NM$ matrix $[\mathbf{D}_1 \ \mathbf{D}_2 \ \cdots \ \mathbf{D}_M]$. An appropriate modification of (42) is easily deduced.

Whilst a full study of the performance of the approximate filtering scheme deduced from the procedure above in specific applications is beyond the scope of this paper, we provide the following simple example for the purpose of illustrating the potential utility of the technique.

Example 4.2

Let the state space be $\mathcal{X} = \{1, \dots, N\}$ for some integer $N > 1$, and let the observation space be $\mathcal{Y} = \{1, \dots, N^2\}$. We define the observations according to

$$Y_t = [N(X_{t-1} - 1) + X_t + V_t],$$

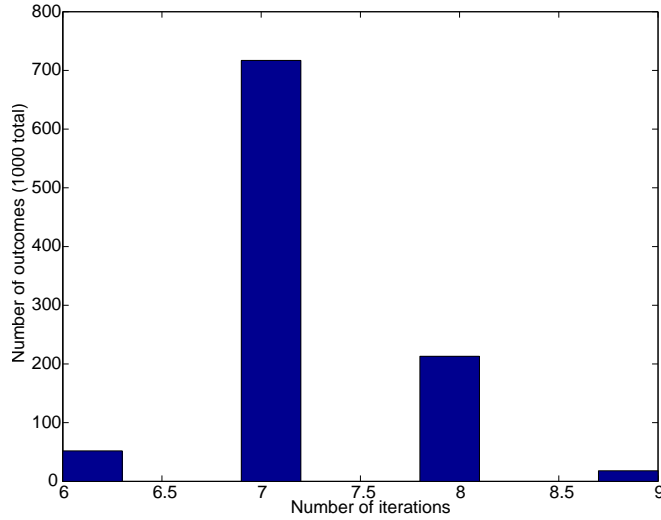


Figure 1: Histogram of the number of iterations of the convex projection algorithm for $N = 100$ and $n = 2$

where V_t is an independent and identically distributed sequence of zero-mean normal random variables with variance σ^2 , and $[\cdot]$ denotes taking the integer part, with appropriate consideration for the boundaries. Thus Y_t is a lexicographic ordering of (X_{t-1}, X_t) with a disturbance increasing with increasing σ^2 due to the V_t . Therefore the conditional probability distribution of Y_t given (X_{t-1}, X_t) is specified. The state transition probabilities were randomly chosen, and the model remained fixed over all experiments. The state aggregation was the two subsets $\{1, \dots, N_1\}$ and $\{N_1 + 1, \dots, N\}$ for some $1 \leq N_1 < N$.

To compare the performance of the optimal filter for the *aggregated* states to the filter derived from the approximate 2-lumpable approximation, we generated 10000 independent realisations of the above FSS, each of length 100 samples. We chose $N = 5$ and $N_1 = 2$. Figure 2 shows the state estimation error probability for each case, as σ^2 varies. As expected, the performance of both filters degrades with increasing σ^2 , the approximate filter performs well and remains surprisingly robust.

5 Conclusions

This paper generalises the strong lumpability results for hidden Markov chains (HMCs) of [21] in a number of significant ways. It has established the theory for the strong lumpa-

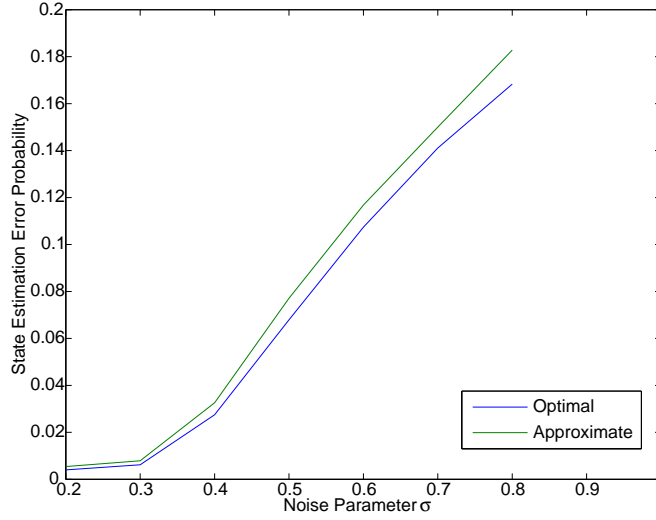


Figure 2: Performance of the Optimal and Approximate filters

bility of finite stochastic systems (FSSs) which include HMCs as a special case. A (weak) lumpability condition based on [12, 16] is introduced and discussed for FSSs. The main results are on the linear dynamics of the lumped FSS. First, a necessary and sufficient condition for the dynamics of the lumped process to be linear was established. Second, using the (weak) lumpability condition, the filter for the FSS can be directly derived from the filter for the lumped FSS (and the lumpability condition is shown to be necessary for this specific derivation to hold). Finally, an algorithm for approximation of an arbitrary FSS, by a strongly lumpable FSS was proposed.

A Proof that statements (d) and (f) in Lemma 2.1 are equivalent for any $n > 2$

We assume right singular vectors of \mathbf{L} are ordered so that $\text{Ker}(\mathbf{L}) = \text{Span}\{\mathbf{v}_{n+1}, \dots, \mathbf{v}_N\}$. The (i, j) element of the matrix $\mathbf{V}\mathbf{Q}\mathbf{V}^\top$ is $\mathbf{v}_i\mathbf{Q}\mathbf{v}_j^\top$. Let $i \in \{n+1, \dots, N\}$ and $j \in \{1, \dots, n\}$, and consider $\mathbf{v}_i\mathbf{Q} \in \text{Ker}(\mathbf{L}) \mathbf{Q} \subset \text{Ker}(\mathbf{L})$ (by assumption (d) in Lemma 2.1). Thus $\mathbf{v}_i\mathbf{Q} \perp \mathbf{v}_j$ since the right singular vectors are orthogonal ([8], Thrm. 2.5.2, p. 70). Hence the lower left hand block of $\mathbf{V}\mathbf{Q}\mathbf{V}^\top$ as shown in (13) is zero as claimed. Now the range of \mathbf{L}^\top is spanned by the rows of \mathbf{L} , and these are precisely the usual unit orthonormal basis. Thus there is an ordering of the \mathbf{v}_i , consistent with the above ordering such

that the matrix of left singular vectors of \mathbf{L}^\top is the identity matrix of size n . Thus we can write the singular value decomposition [8] $\mathbf{L}^\top = \mathbf{S} \mathbf{V}$, where $\mathbf{S} = [\mathbf{S}_1 \ 0]$ with \mathbf{S}_1 is an $n \times n$ diagonal matrix with strictly positive entries. Let $\mathbf{V}^\top = [\mathbf{V}_1^\top \ \mathbf{V}_2^\top]$ where \mathbf{V}_1^\top is of size $N \times n$. Consider

$$\begin{aligned} \mathbf{L} (\mathbf{L}^\top \mathbf{L})^{-1/2} &= \mathbf{V} \mathbf{S}^\top (\mathbf{S} \mathbf{S}^\top)^{-1/2} \\ &= \mathbf{V}_1 \mathbf{S}_1 (\mathbf{S}_1 \mathbf{S}_1)^{-1/2} \\ &= \mathbf{V}_1. \end{aligned}$$

The matrix on the left is non-negative, thus so is \mathbf{V}_1 . So, in (13), the matrix $\mathbf{B}_{11} = \mathbf{V}_1 \mathbf{Q} \mathbf{V}_1^\top$ is non-negative (since \mathbf{Q} has non-negative entries).

Now let $\mathbf{x} \in \text{Ker}(\mathbf{L})\mathbf{Q}$, then there are scalars $\alpha_{n+1}, \dots, \alpha_N$ such that

$$\mathbf{x} = \sum_{i=n+1}^N \alpha_i \mathbf{v}_i \mathbf{Q}.$$

Let $j \in \{1, \dots, n\}$, and consider

$$\begin{aligned} \mathbf{x} \mathbf{v}_j^\top &= \sum_{i=n+1}^N \alpha_i \mathbf{v}_i \mathbf{Q} \mathbf{v}_j^\top \\ &= 0, \end{aligned}$$

by Lemma 2.1 (f). Thus $\mathbf{x} \perp \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\} = \text{Ker}(\mathbf{L})^\perp$, so $\mathbf{x} \in \text{Ker}(\mathbf{L})$ establishing Lemma 2.1 (d).

B Direct proof that statements (a)–(f) in Theorem 3.1 are equivalent

First, we know that the conditional probability in (20) is independent of y_0 . It follows from (1) that they are equal to

$$\mathbb{P}\{Y_{t+1} = y_1, g(X_{t+1}) = w_2 \mid Y_t = y_0, X_t = x_1\} = \sum_{x_2 \in g^{-1}(w_2)} D_{y_1}(x_1, x_2).$$

This is exactly what is required in (a). Condition (a) can be reformulated in a matrix form as: for every $y_1 \in \mathcal{Y}$,

$$\mathbf{D}_{y_1} \mathbf{L} = \mathbf{L} \hat{\mathbf{D}}_{y_1} \tag{43}$$

where \hat{D}_{y_1} is a non-negative $n \times n$ -matrix, and $\hat{D}_{y_1}(w_1, w_2)$ is the common value of the probabilities in (15). Then, using $UL = I_n$, we obtain (17). By expressing the matrix equality (43) element by element and using (1), condition (a) is obtained. Therefore, conditions (a) and (c) are equivalent.

Next assume that (c) holds. Let $x \in \text{Ker}(L)$, then it follows from (17) that

$$xD_yL = xLUD_yL = 0.$$

Therefore, $xD_y \in \text{Ker}(L)$ and the condition (d) is satisfied.

If (d) holds, then for any $x \in \text{Ker}(L)$ we have $xD_y \in \text{Ker}(L)$ that is $xD_yL = 0$. Therefore $x \in D_y \text{Ker}(L)$ and (e) is valid. Next, statement (e), i.e. $\text{Ker}(L) \subseteq \text{Ker}(D_yL)$, implies that there exists a $n \times n$ matrix \hat{D} such that $D_yL = L\hat{D}$ (e.g. see [22]). If we multiply from the left the previous relation by U , we obtain (17) and condition (c) is satisfied. Thus, conditions (c),(d),(e) are equivalent.

Note that (16a)-(16b) in statement (b) are just a rewriting of (15) which gives insight in the structure of the matrix D_y required for having strong lumpability. That $\sum_{y \in \mathcal{Y}} \sum_{j=1}^n c_{k,j}(y) = 1$ for each $k = 1, \dots, n$ follows from the fact that $\sum_y D_y$ is a stochastic matrix.

Finally, equivalence of statements (d) and (f) can be deduced as in Appendix A for the MC case.

Acknowledgment The authors thank the referee for his/her comments which led to a substantial improvement of the presentation of the paper.

References

- [1] BARAS, J. S. AND FINESSO, L. (1992). Consistent estimation of the order of hidden Markov chains. In *Stochastic Theory and Adaptive Control*. ed. Duncan, T.E. and Pasik-Duncan, B. vol. 184 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin. pp. 26–39.
- [2] BENOIT, A., BRENNER, L., FERNADES, P. AND PLATEAU, B. (2004). Aggregation of stochastic automata networks with replicas. *Linear Algebra Appl.* **386**, 111–136.
- [3] BREUER, L. AND BAUM, D. (2005). *An Introduction to Queuing Theory*. Springer, Berlin.
- [4] BUCHHOLZ, P. (1994). Exact and ordinary lumpability in finite Markov chain. *J. Appl. Probab.* **31**, 59–75.
- [5] DERISAVI, S., HERMANN, H. AND SANDERS, W. H. (2003). Optimal state-space lumping in Markov chains. *Inform. Process. Lett.* **87**, 309–315.
- [6] EPHRAIM, Y. AND MERHAV, N. (2002). Hidden Markov processes. *IEEE Trans. Inform. Theory* **48**, 1518–1569.

- [7] GILMORE, S., HILLSTON, J. AND RIBAUDO, M. (2001). An efficient algorithm for aggregating PEPA models. *IEEE Trans. Software Eng.* **27**, 449–464.
- [8] GOLUB, G. H. AND VAN LOAN, C. F. (1996). *Matrix Computations*. Ed. 3. Johns Hopkins University Press, Baltimore.
- [9] GRINFELD, M. AND KNIGHT, P. A. (2000). Weak lumpability in the k -sat problem. *App. Math. Lett.* **13**, 49–53.
- [10] GURVITS, L. N. AND LEDOUX, J. (2005). Markov property for a function of a Markov chain: a linear algebra approach. *Linear Algebra Appl.* **404C**, 85–117.
- [11] HERMANN, H., HERZOG, U. AND KATOEN, J.-P. (2002). Process algebra for performance evaluation. *Theoretical Computer Science* **274**, 43–87.
- [12] KEMENY, J. G. AND SNELL, J. L. (1976). *Finite Markov chains*. Springer-Verlag, New York.
- [13] LEDOUX, J. (2005). Recursive filters for partially observable finite Markov chains. *J. Appl. Probab.* **42**, 687–697.
- [14] PAIGE, R. L. (2001). Classification and lumpability in the stochastic hopfield model. *Adv. in Appl. Probab.* **33**, 930–943.
- [15] PICCI, G. (1978). On the internal structure of finite state stochastic processes. In *Recent developments in Variable Structure Systems*. ed. Mohler R.R and Ruberti, A. vol. 162 of *Lecture Notes in Economics and Mathematical systems*. Springer, Berlin pp. 288–304.
- [16] ROGERS, L. C. G. AND PITMAN, J. W. (1981). Markov functions. *Ann. Probab.* **9**, 573–582.
- [17] SPREIJ, P. J. C. (2001). On the Markov property of a finite hidden Markov chain. *Statist. Probab. Lett.* **52**, 279–288.
- [18] SPREIJ, P. J. C. (2003). On hidden Markov chains and finite stochastic systems. *Statist. Probab. Lett.* **62**, 189–201.
- [19] SPROSTON, J. AND DONATELLI, S. (2006). Backward bisimulation in Markov chain model checking. *IEEE Trans. Software Eng.* **32**, 531–546.
- [20] TIAN, J. AND LIN, X.-S. (2005). Colored coalescent theory. *Discrete Contin. Dyn. Syst.* 833–845.
- [21] WHITE, L. B., MAHONY, R. AND BRUSHE, G. D. (2000). Lumpable hidden Markov models—model reduction and reduced complexity filtering. *IEEE Trans. Automat. Control* **45**, 2297–2306.
- [22] WONHAM, W. M. (1979). *Linear Multivariable Control: a Geometric Approach*. Springer-Verlag, New York.
- [23] YOULA, D. C. (1987). Mathematical theory of image restoration by the method of convex projections. In *Image Recovery : Theory and Applications*. ed. H. Stark. Academic Press, New York.